(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
30 May 2003 (30.05.2003)

PCT

(10) International Publication Number
WO 03/044194 A1

(51) International Patent Classification⁷: C12N 15/10, C12P 21/00, C12Q 1/68, G01N 33/53, 33/566, C07H 21/02, C08H 1/00

(21) International Application Number: PCT/US02/37103

(22) International Filing Date:
18 November 2002 (18.11.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/346,965 16 November 2001 (16.11.2001) US

(71) Applicant *(for all designated States except US)*: PROTEONOVA, INC. [US/US]; 814 South Euclid Ave., Pasadena, CA 91106 (US).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: WILLIAMS, Richard, B. [US/US]; 1104 Orange Grove Ave., South Pasadena, CA 91030 (US).

(74) Agent: ALTMAN, Daniel, E.; Knobbe, Martens, Olson and Bear, LLP, 2040 Main Street, Fourteenth Floor, Irvine, CA 92614 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT (utility model), AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ (utility model), CZ, DE (utility model), DE, DK (utility model), DK, DM, DZ, EC, EE (utility model), EE, ES, FI (utility model), FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SI, SK (utility model), SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declaration under Rule 4.17:
— *of inventorship (Rule 4.17(iv)) for US only*

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: SELECTION AND EVOLUTION OF NUCLEIC ACIDS AND POLYPEPTIDES

(57) Abstract: This invention relates to methods and reagents for selecting a desired protein or nucleic acid molecule by linking mRNA, with known or unknown sequences, to its translated protein to form a cognate pair. The cognate pair is selected based upon desired properties of the protein or the nucleic acid. This method also includes the evolution of a desired protein or nucleic acid molecule by amplifying the nucleic acid portion of the selected cognate pair, introducing variation into the nucleic acid, translating the nucleic acid, attaching the nucleic acid to its protein to form a second cognate pair, and re-selecting this cognate pair based upon desired properties.

WO 03/044194 A1

SELECTION AND EVOLUTION OF NUCLEIC ACIDS AND POLYPEPTIDES

Field of the Invention

The present invention relates to compositions and methods for the selection of nucleic acids and polypeptides.

Background of the Invention

Ligand-receptor interactions are of interest for many reasons, from elucidating basic biological site recognition mechanisms to drug screening and rational drug design. It has been possible for many years to drive *in vitro* evolution of nucleic acids by selecting molecules out of large populations that preferentially bind to a selected target, then amplifying and mutating them for subsequent re-selection (Tuerk and Gold, Science 249:505 (1990), herein incorporated by reference).

The ability to perform such a selection process with proteins would be extremely useful. This would permit *in vitro* design and production of proteins that bind specifically to chosen ligands. The use of proteins, as compared to nucleic acids, is particularly advantageous because the twenty diverse amino acid side chains in proteins have far more binding possibilities than the four similar chains in nucleic acid side. Further, many biologically and medically relevant ligands bind proteins.

Both nucleic acid and protein evolution methods require access to a large and highly varied population of test molecules, a way to select members of the population that exhibit the desired properties, and the ability to reproduce the selected molecules with mutated variations to obtain another large population for subsequent selection.

Prior attempts to develop a protein evolution method were primarily limited by the inability of the proteins to reproduce themselves and the inability to link mRNA encoding a polypeptide with the translated product. Additionally, the generation of large peptide libraries and screening methods have, until recently, required that the process have an *in vivo* expression step. Examples include yeast two- or three-hybrid, yeast display and phage display methods (Fields and Song, Nature 340:245 (1989); Licitra and Liu, PNAS 93:12817 (1996); Boder and Wittrup, Nat Biotechnol 15:553 (1997); and Scott and Smith,

Science 249:386 (1990)). *In vivo* methods suffer from various disadvantages, including a limited library size and cumbersome screening steps. Additionally, undesirable selective pressures can be placed on the generation of variants by cellular constraints of the host.

In vitro methods have been developed more recently, using prokaryotic and eukaryotic *in vitro* translation systems, such as ribosome display (Mattheakis et al., PNAS 91:9022 (1994); Hanes and Plückthun, PNAS 94:4937 (1997); Jermutus et al., Current Opinion in Biotechnology 9:534 (1998), all herein incorporated by reference). These methods link the protein and its encoding mRNA with the ribosome, and the entire complex is screened against a ligand of choice. Potential disadvantages of this method include the large size of the ribosome, which could interfere with the screening of the attached, and relatively tiny, protein.

In 1997, two groups of workers developed an *in vitro* method of attaching a protein to its coding sequence during translation by using the ribosomal peptidyl transferase with puromycin attached to a linker DNA (Szostak et al., International Patent Publication WO 98/31700; Roberts and Szostak PNAS 94:12297 (1997); Nemoto et al., FEBS Letters 414:405 (1997), all herein incorporated by reference). Once the coding sequence and peptides are linked, the peptides are exposed to a selected ligand. Selection or binding of the peptide by the ligand also selects the attached coding sequence, which can then be reproduced by standard means. Both Roberts and Szostak and Nemoto et al. used the technique of attaching a puromycin molecule to the 3' end of a coding sequence by a DNA linker or other non-translatable chain. Puromycin is a tRNA acceptor stem analog which accepts the nascent peptide chain under the action of the ribosomal peptidyl transferase and binds it stably and irreversibly, thereby halting translation. These methods suffer from certain disadvantages. For example, the coding sequence encoding each peptide must be known and be modified both initially and between each selection. Thus, the methods of Roberts and Nemoto cannot be used to select native unknown mRNAs. Further, the modification of the coding sequence adds several steps to the process. Finally, the attached puromycin on the linker molecules may compete in the translation reaction with the native tRNAs for the A site on the ribosome reading its coding sequence or a nearby ribosome, and could thus "poison" the translation process, just as would unattached puromycin in the translation reaction solution. Inadvertent interactions between puromycin and ribosomes could result in two kinds of reaction non-specificity: prematurely shortened proteins and

proteins attached to the wrong message. There are reports in the prior art that indicate that the avidity of the A site and the peptidyl transferase for the puromycin may be modulated by Mg^{++} concentration (Roberts, Curr. Opin. Chem. Biol. 3:268 (1999), herein incorporated by reference). Although Mg^{++} concentration may be titrated to control for the first kind of non-specificity, i.e. premature termination of translation, it will not affect the second type, i.e. inaccurate mRNA-protein linkage.

Thus, a need exists for an *in vitro* nucleic acid-based protein evolution system that does not necessarily require initial knowledge of the nucleic acid's sequence or repeated chemical modification of the nucleic acids, and which can accurately link a mRNA to its protein.

Summary of the Invention

The present invention provides compositions and methods to select and evolve desired properties of proteins and nucleic acids. In various embodiments, the current invention provides modified tRNA's and tRNA analogs. Other embodiments include methods for generating polypeptides, assays enabling selection of individual members of the population of polypeptides having desired characteristics, methods for amplifying the nucleic acids encoding such selected polypeptides, and methods for generating new variants to screen for enhanced properties.

In several embodiments, the present invention permits the attachment of a protein to its respective mRNA without requiring modification of native mRNA, although modified mRNA may still be used. The specificity of the methods embodied in various aspects of the current invention are determined by the specificity of the codon-anticodon interaction.

In a preferred embodiment, the invention permits the selection of nucleic acids by selecting the proteins for which they code. This may be accomplished by connecting the protein to its cognate mRNA at the end of translation, which in turn is done by connecting both the protein and mRNA to a tRNA or tRNA analog.

A preferred embodiment of the invention includes a tRNA molecule capable of covalently linking a nucleic acid encoding a polypeptide and the polypeptide to the tRNA, wherein the linkage of the nucleic acid occurs on a portion of the tRNA other than the linkage to the polypeptide and wherein the tRNA comprises a linking molecule associated with the anticodon of the tRNA. This anticodon of the tRNA is capable of forming a

crosslink to the mRNA under irradiation with light of a required wavelength, preferably a furan-sided psoralen monoadduct on the anticodon irradiated with UVA, preferably in the range of about 300-450 nm, more preferably in the range of about 320 to 400 nm, and most preferably about 365 nm. Preferably, an amino acid or amino acid analog is attached to the 3' end of a tRNA molecule by a stable bond to generate a stable aminoacyl tRNA analog (SATA).

Other embodiments include a mRNA comprising a psoralen, preferably located in the 3' region of the reading frame, more preferably at the most 3' codon of the reading frame, most preferably at the 3' stop codon of the reading frame. In preferred embodiments, linkage between the tRNA and the mRNA is a cross-linked psoralen molecule, more preferably a furan-sided psoralen monoadduct.

A further embodiment of the invention provides a method of forming a monoadduct. According to this method a target oligonucleotide with at least one uridine and at least one modified uridine is contacted with psoralen, and the target oligmeleotide and psoralen are coupled to form a monoadduct. The modified uridine according to this embodiment may be modified to avoid coupling with psoralen, and preferably the modified uridine is pseudouridine. According to this embodiment the target oligonucleotide may be a tRNA molecule, such as tRNA, modified tRNA and tRNA analogs or a mRNA molecule, such as mRNA, modified mRNA and mRNA analogs. In a further embodiment the psoralen is coupled to the target oligonucleotide by one or more cross-links. According to this embodiment a second oligonucleotide with a nucleotide sequence complementary to the target oligonucleotide sequence may be present. This second oligonucleotide may contain no uridine or may contain uridine residues that are modified to avoid cross-linking with the target oligonucleotide. Preferably, the modified uridine is pseudouridine.

Several embodiments of the present invention include a method of stably linking a nucleic acid, a tRNA, and a polypeptide encoded by the nucleic acid together to form a linked nucleotide-polypeptide complex. In a preferred embodiment, the nucleic acid is an mRNA and the linked nucleotide-polypeptide complex is a mRNA-polypeptide complex. The method can further comprise providing a plurality of distinct nucleic acid-polypeptide complexes, providing a ligand with a desired binding characteristic, contacting the complexes with the ligand, removing unbound complexes, and recovering complexes bound to the ligand.

Several methods of the current invention involve the evolution of nucleic acid molecules and/or proteins. In one embodiment, this invention comprises amplifying the nucleic acid component of the recovered complexes and introducing variation to the sequence of the nucleic acids. In other embodiments, the method further comprises translating polypeptides from the amplified and varied nucleic acids, linking them together using tRNA, and contacting them with the ligand to select another new population of bound complexes. Several embodiments of the present invention use selected protein-mRNA complexes in a process of *in vitro* evolution, in particular the iterative process in which the selected mRNA is reproduced with variation, translated and again connected to cognate protein for selection.

Brief Description of the Drawings

Figure 1 illustrates schematically one example of the complex formed by the mRNA and its protein product when linked by a modified tRNA or analog. As shown, a codon of the mRNA pairs with the anticodon of a modified tRNA and is covalently crosslinked to a psoralen monoadduct by UV irradiation. The translated polypeptide is linked to the modified tRNA via the ribosomal peptidyl transferase. Both linkages occur while the mRNA and nascent protein are held in place by the ribosome.

Figure 2 illustrates schematically an example of the *in vitro* selection and evolution process, wherein the starting nucleic acids and their protein products are linked (e.g., according to Figure 1) and are selected by a particular characteristic exhibited by the protein. Proteins not exhibiting the particular characteristic are discarded and those having the characteristic are amplified with variation, preferably via amplification with variation of the mRNA, to form a new population. In various embodiments, nonbinding proteins will be selected. The new population is translated and linked via a modified tRNA or analog, and the selection process is repeated. As many selection and amplification/mutation rounds as desired can be performed to optimize the protein product.

Figure 3 illustrates one method of construction of a tRNA molecule of the invention. In this embodiment, the 5' end of a tRNA, a nucleic acid encoding an anticodon loop and having a molecule capable of stably linking to mRNA (such as psoralen, as used in this example), and the 3' end of tRNA modified with a terminal puromycin molecule are

ligated to form a complete modified tRNA for use in the *in vitro* evolution methods of the invention.

Figure 4 describes two alternative embodiments by which the crosslinking molecule psoralen can be positioned such that it is capable of linking the mRNA with the tRNA in the methods of the invention. A first embodiment includes linking the psoralen monoadduct to the mRNA, and a second embodiment includes linking the psoralen to the anticodon of the tRNA. Psoralen can either be monoadducted to the anticodon or the 3' terminal codon of the reading frame for known or partially known messages. This can be done in a separate procedure from translation, i.e. before translation occurs.

Figure 5 illustrates the chemical structures for uridine and pseudouridine. Pseudouridine is a naturally occurring base found in tRNA that forms hydrogen bonds just as uridine does, but lacks the 5-6 double bond that is the target for psoralen.

Detailed Description of the Preferred Embodiment

Various aspects of the present invention use a tRNA mechanism that links messenger RNA (mRNA) to its translated protein product, forming a "cognate pair." In several embodiments, mRNA whose sequence is not known can be expressed, its protein characterized through a selection process against a ligand with desired or selected properties, and nucleic acid evolution - resulting in protein evolution- can be performed *in vitro* to arrive at molecules with enhanced properties. The cognate pairs are preferably attached via a linking tRNA, modified tRNA, or tRNA analog. In a preferred embodiment, the tRNA is connected to the nascent peptide by the ribosomal peptidyl transferase and to the mRNA through an ultraviolet induced cross link between the anticodon of the tRNA or tRNA analog and the codon of the RNA message. This can be done by, for example thiouracil, but in a preferred method, the linker is a psoralen crosslink made from a psoralen monoadduct pre-placed on either the mRNA or preferably on the tRNA anticodon of choice. Preferably, a tRNA stop anticodon is selected. A stop codon/anticodon pair selects for full length transcripts. One skilled in the art will understand that a mRNA not having a stop codon may also be used and that any codon or nucleic acid triplet may be used. A tRNA having an anticodon which is not naturally occurring can be synthesized according to methods known in the art (e.g. Figure 3).

The terms "protein," "peptide," and "polypeptide" are defined herein to mean a polymeric molecule of two or more units comprised of amino acids in any form (e.g., D- or L- amino acids, synthetic or modified amino acids capable of polymerizing via peptide bonds, etc.), and these terms may be used interchangeably herein.

The term "pseudo stop codon" is defined herein to mean a codon which, while not naturally a nonsense codon, prevents a message from being further translated. A pseudo stop codon may be created by using a "stable aminoacyl tRNA analog" or SATA, as described below. In this manner, a pseudo stop codon is a codon which is recognized by and binds to a SATA. Another method by which to create a pseudo stop codon is to create an artificial system in which the necessary tRNA having an anticodon complementary to the pseudocodon is substantially depleted. Accordingly, translation will stop when the absent tRNA is required, i.e. at the pseudo stop codon. One skilled in the art will appreciate that are numerous ways to create a pseudo stop codon as defined herein.

The formation of connections between mRNA and its protein product generally requires a tRNA or tRNA analog with certain characteristics. In several embodiments of the current invention, the tRNA or tRNA analog will have a stable peptide acceptor. This modification changes the tRNA or tRNA analog such that after it accepts the nascent peptide chain by the action of the ribosomal peptidyl transferase, it holds the chain in a stable manner such that the peptidyl transferase cannot detach it. This may be accomplished by using a bond such as a 2' ester on a 3' deoxy adenosine or an amino "acyl $tRNA_{ox\text{-}red}$" which can bind to the ribosome, accept the peptide chain, and then not act as a donor in the next transpeptidation (Chinali et al., Biochem. 13:3001 (1974); Krayevsky and Kukhanova, Prog. Nuc. Acid Res 23:1 (1979) and Sprinzl and Cramer Prog. Nuc. Acid Res 22:1 (1979), all herein incorporated by reference).

In a preferred embodiment, an amino acid or amino acid analog is attached to the 3' end of the tRNA or tRNA analog by a stable bond. This stable bond contrasts the labile, high energy ester bond that connects these two in the native structure. The stable bond not only protects the bond from the action of the peptidyl transferase, but also preserves the structure during subsequent steps. For convenience, this modified tRNA or tRNA analog will be referred to as a "stable aminoacyl tRNA analog" or SATA. As used herein, a SATA is an entity which can recognize a selected codon such that it can accept a peptide chain by the action of the ribosomal peptidyl transferase when the cognate codon is in the reading

position of the ribosome. The peptide chain will be bound in such a way that the peptide is bound stably and cannot be unattached by the peptidyl transferase. Preferably, the selected codon is recognized by hydrogen bonding.

One method for creating a SATA was published in 1973 (Fraser and Rich, PNAS 70:2671 (1973), herein incorporated by reference). This method involves the conversion of a tRNA, or tRNA analog, to a 3'-amino-3'-deoxy tRNA. This is accomplished by adding a 3'-amino-3'-deoxy adenosine to the end of a native tRNA with tRNA nucleotidyl transferase after removing the native adenosine from it with snake venom phosphodiesterase. This modified tRNA is then charged with an amino acid by the respective aminoacyl tRNA synthetase (aaRS). Fraser and Rich used an aaRS in which the tRNA is charged on the 3', rather than the 2', hydroxyl. The amino acid is bound to the tRNA by a stable amide bond rather than the usual labile high-energy ester bond. Thus, when it accepts a peptide from ribosomal peptidyl transferase it will stably hold the peptide and not be able to donate it to another acceptor.

In a preferred method, the SATA will be attached to the translated message by a psoralen cross link between the codon and anticodon. Psoralen cross links are preferentially made between sequences that contain complementary 5' pyrimidine-purine 3' sequences especially UA or TA sequences (Cimino et al., Ann. Rev. Biochem. 54:1151 (1985), herein incorporated by reference). The codon coding for the SATA, or the linking codon, can be PYR-PUR-X or X-PYR-PUR, so that several codons may be used for the linking codon. Conveniently, the stop or nonsense codons have this configuration. Using a codon that codes for an amino acid may require minor adjustments to the genetic code, which could complicate some applications. Therefore, in a preferred embodiment, a stop codon is used as the linking codon and the SATA functions as a nonsense suppressor in that it recognizes the linking codon. One skilled in the art, however, will appreciate that, with appropriate adjustments to the system, any codon can be used.

Fraser and Rich did their work in *E. coli*, but the most effective *in vitro* translation systems are in eukaryotes The use of prokaryotic suppressors in eukaryotic translation systems appears to be feasible (Geller and Rich Nature 283:41 (1980); Edwards et al PNAS 88:1153 (1991); Hou and Schimmel Biochem 28:6800 (1989), all herein incorporated by reference). They are primarily limited by the resident aaRS's. This limitation is overcome by various embodiments of the present invention because the tRNA or analog can be

charged in the prokaryotic system and then purified according to established methods (Lucas-Lenard and Haenni, PNAS 63:93 (1969), herein incorporated by reference).

In several embodiments of the current invention, acceptor stem modifications suitable for use in the tRNAs and analogs can be produced by various methods known in the art. Such methods are found in, for example, Sprinzl and Cramer, Prog. Nuc. Acid Res. 22:1 (1979), herein incorporated by reference. In an alternative embodiment, "transcriptional tRNA", i.e. the sequence of the tRNA as it would be transcribed rather than after the post-transcriptional processing, leads to the atypical and modified bases that are common in tRNAs. These transcriptional tRNAs are capable of functioning as tRNAs (Dabrowski et al., EMBO J. 14: 4872, 1995; and Harrington et al., Biochem. 32: 7617, 1993, both herein incorporated by reference). Transcriptional tRNA can be produced by transcription or can be made by connecting commercial RNA sequences together, piece-wise as in Figure 3, or in some combination of established methods. For instance, the 5' phosphate and 3' puromycin are commercially available attached to oligoribonucleotides. Commercial RNA sequences are available from Dharmacon Research Inc., La Fayette, CO. This company can also provide modified native tRNA, such as sequences in which thymine is substituted for uricil and pseudouridine.) These pieces can be connected together using T4 DNA ligase, as is well-known in the art (Moore and Sharp, Science 256: 992, 1992, herein incorporated by reference). Alternatively, in a preferred embodiment, T4 RNA ligase is used (Romaniuk and Uhlenbeck, Methods in Enzymology 100:52 (1983), herein incorporated by reference).

In several embodiments of the present invention, psoralen is monoadducted to the SATA by construction of a tRNA from pieces including a psoralen linked oligonucleotide (Fig. 3) or by monoadduction to a native or modified tRNA or analog (Fig. 4). In a preferred embodiment, psoralen is first monoadducted to an oligonucleotide containing part of the anticodon loop as described below and this product is then ligated to the remaining fragments of the SATA.

In several embodiments, translation will stop when the nascent protein is attached to the SATA by the peptidyl transferase. When a large number of ribosomes are in this position the SATA and the mRNA will be connected with UV light. In a preferred method this will be accomplished by having a psoralen crosslink formed. Psoralens have a furan side and a pyrone side, and they readily intercalate between complementary base pairs in

double stranded DNA, RNA, and DNA-RNA hybrids (Cimino et al., Ann. Rev. Biochem. 54:1151 (1985), herein incorporated by reference). Upon irradiation with UV, preferably in the range of 320 nm to 400 nm, cross linking will take place and leave the staggered pyrimidines covalently bound. By either forming crosslinks and photo reversing them or by using selected wavelengths, it is possible to form monoadducts, described more fully below. These will be either pyrone sided or furan sided monoadducts. Upon further irradiation, the furan sided monoadducts can be covalently crosslinked to complementary base pairs. The pyrone sided monoadducts cannot be further crosslinked. The formation of the furan sided psoralen monoadduct (MAf) is also done according to established methods. In a preferred method, the psoralen is attached to the anticodon of the SATA. However, psoralen can also be attached at the end of the reading frame of the message, as depicted in Figure 4.

Methods for large scale production of purified MAf on oligonucleotides are described in the literature (e.g., Speilmann et al., PNAS 89:4514, 1992, herein incorporated by reference), as are methods that require less resources, but have some non-cross-linkable pyrone sided psoralen monoadduct contamination (e.g., U.S. Patent No. 4,599,303; Gamper et al., J. Mol. Biol. 197: 349 (1987); Gamper et al., Photochem. Photobiol. 40:29 (1984), both herein incorporated by reference). In several embodiments of the current invention, psoralen labeling is accomplished by using either method. In a preferred embodiment, furan sided monoadducts will be created using visible light, preferably in the range of approximately 400 nm - 420 nm, according to the methods described in U.S. Patent No. 5,462,733 and Gasparro et al., Photochem. Photobiol. 57:1007 (1993), both herein incorporated by reference. In one aspect of this invention, a SATA with a furan sided monoadduct or monoadducted oligonucleotides for placement on the 3' end of mRNAs, along with a nonadducted SATA are provided as the basis of a kit.

In one embodiment, the formation and reversal of monoadducts and crosslinks are performed according to the methods of Bachellerie et al. (Nuc Acids Res 9:2207 (1981)), herein incorporated by reference. In a preferred embodiment, efficient production of monoadducts, resulting in high yield of the end-product, is accomplished using the methods of Kobertz and Essigmann, J. A. Chem. Soc. 1997, 119, 5960-5961 and Kobertz and Essigmann, J. Org. Chem. 1997, 62, 2630-2632, both herein incorporated by reference.

In a preferred embodiment, a SATA fragment and complementary RNA or DNA is used in which all of the uridines, except the target, are replaced by pseudouridine. Figure 5 compares the chemical structures for uridine and pseudouridine. Pseudouridine is a naturally occurring base found in tRNA that forms hydrogen bonds just as uridine does. This embodiment is particularly advantageous because the pseudouridine forms the same Watson-Crick hydrogen-bonds as the native uridine but lacks the 5-6 double bond that is the target for interacting with either the furan or pyrone side of the psoralen molecule. This permits the same base-pairing characteristics as an oligonucleotide with uridine, but provides only one target for the psoralen. Because the pyrone side linkage is usually formed after the furan side has reacted, this removal of a staggered target allows the monoadduct to be formed with high efficiency irradiation without forming crosslinks and with minimal formation of pyrone sided monoadduct (MaP). Irradiation is preferably in the range of about 300-450 nm, more preferably in the range of about 320 to 400 nm, and most preferably about 365 nm. More specifically, a pseudouridine on the SATA permits: 1) the use of SATA sequences that contain uridines which are potential targets for the psoralen and 2) on the cRNA or cDNA, eliminate the formation of crosslinks, leaving the process stopped at furan sided monoadduct (MaF) formation when using UVA wavelengths which are much more efficient than visible light. Use of the SATA and the monoadduct in several embodiments of the current invention is particularly advantageous for *in vitro* translation systems. However, one skilled in the art will appreciate that in situ systems can also be used. Various embodiments of the current invention will be applicable to any *in vitro* translation system, including, but not limited to, rabbit reticulocyte lysate (RLL), wheat germ, *E. coli*, and yeast lysate systems. Many embodiments of the current invention are also well-suited for use in hybrid systems where components of different systems are combined.

tRNAs aminoacylated on a 3' amide bond are reported not to combine with the elongation factor EF-TU which assists in binding to the A site (Sprinzl and Cramer, Prog. Nuc. Acid Res. 22:1 (1979), herein incorporated by reference). Such modified tRNAs do, however, bind to the A site. This binding of 3' modified tRNAs can be increased by changing the Mg^{++} concentration (Chinali et al., Biochem. 13:3001 (1974), herein incorporated by reference). The appropriate concentrations and/or molar ratios of SATA and Mg^{++} can be determined empirically. If the concentration or A site avidity of SATA is

too high, the SATA could compete with native tRNAs for non-cognate codons i.e., could function much like puromycin and stall translation. If the concentration or A site avidity of SATA is too low, the SATA might not effectively compete with the release factors, i.e., it would not act as an effective nonsense suppressor tRNA. The balance between these can be determined empirically.

It is also believed that the elongation factor aids in proofreading the codon-anticodon recognition. The error rate in the absence of elongation factor and the associated GTP hydrolysis is estimated to be 1 in 100 for codons one nucleotide away (Voet and Voet, Biochemistry 2nd ed. pp. 1000-1002 (1995), John Wiley and Sons, herein incorporated by reference). In a preferred embodiment, UAA is used as the linking codon. For UAA as the linking codon, there are 7 non stop codons which differ by one amino acid. This is 7/61 or about 11.5% of the non stop codons. One can estimate the probability of miscoding a given codon as $(0.01)(0.115) = 1.15 \times 10^{-3}$ miscodes per codon. Thus, one would expect a miscode about every 870 codons, a frequency which will not substantially impair performance of various methods of the current invention. In an alternative embodiment, UAGor UGA is used as the linking codon.

In several embodiments, appropriate concentrations of SATA and Mg^{++} are used in the *in vitro* translation system, e.g. RRL, in the presence of the mRNA molecules in the pool, causing translation to cease when the ribosome reaches the codon which permits the SATA to accept the peptide chain (the linking codon described above). Within a short time, most of the linking codons will be occupied by SATAs within ribosomes. In a preferred embodiment, the system then will be irradiated with UV light, preferably at approximately 320 nm to 400 nm. Nucleic acids are typically transparent to, i.e. do not absorb, this wavelength range. Upon irradiation, the psoralen monoadduct will convert to a crosslink connecting the anticodon and the codon by a stable covalent bond.

In a preferred embodiment, the target mRNA is pre-selected. In another embodiment, the target mRNA is artificially produced. In an alternative embodiment, the target consists of messages native to the system under investigation, which may be unknown and/or unidentified. The ability to use unknown and/or unidentified mRNAs is a particular advantage of several embodiments of the current invention.

In several embodiments, once all the nascent proteins are connected to their cognate mRNAs, the ribosomes are released or denatured. Preferably, this is accomplished by the

depletion of Mg^{++} through dialysis, simple dilution, or chelation. One skilled in the art will understand that other methods, including, but not limited to, denaturation by changing the ionic strength, the pH, or the solvent system can also be used.

In several embodiments of the invention, the selection of cognate pairs will be based upon affinity binding of proteins according to any of a variety of established methods, including, but not limited to, affinity columns, immunoprecipitation, and many high throughput screening procedures. A variety of ligands may also be used, including, but not limited to, proteins, nucleic acids, chemical compounds, polymers and metals. In addition, cell membranes or receptors, or even entire cells may be used to bind the cognate pairs. The selection can be positive or negative. That is, the selected cognate pairs can be those that do bind well to a ligand or those that do not. For instance, for a protein to accelerate a thermodynamically favorable reaction, i.e., act as an enzyme for that reaction, it should bind both the substrate and a transition state analog. However, the transition state analog should be bound much more tightly than the substrate. This is described by the equation

$$\frac{k_{enzyme}}{k_{genzyme}} = \frac{K_{trans}}{K_{subst}}$$

where the ratio of the rate of the reaction with the enzyme, k_{enzyme}, to the rate without, $k_{genzyme}$, is equal to the ratio of the binding of the transition state to the enzyme K_{trans} over the binding of the substrate to the enzyme K_{subst} (Voet and Voet, Biochemistry 2nd ed. p.380, (1995), John Wiley and Sons, herein incorporated by reference).

In a preferred embodiment, proteins which compete poorly for binding to the substrate but compete well for binding to the transition state analog are selected. Operationally, this may be accomplished by taking the proteins that are easily eluted from a matrix with substrate or substrate analog bound to it and are the most difficult to remove from matrix with transition state analog bound to it. By sequentially repeating this selection and reproducing the proteins through replication and translation of the nucleic acid of the cognate pairs, an improved enzyme should evolve. Affinity to one entity and lack of affinity to another in the same selection process is used in several embodiments of the current invention. Selection can also be done by RNA in many embodiments.

Once the selection has identified a population of cognate pairs it may be convenient to detach the mRNA strand from the SATA to reproduce it. This is not always necessary but can be accomplished by irradiating the pairs with UV, preferably at approximately 313 nm or just below. This has been identified as a wave length that will photoreverse the psoralen crosslink to MAf and damage the nucleic acid minimally. The ratio of photoreversal to nucleic acid damage is estimated to be 1 photoreversal for damage to 1 in 600 bases (Cimino et al., Biochem 25:3013 (1986), herein incorporated by reference).

One skilled in the art will appreciate that the mRNAs can be reproduced in many ways including, but not limited to, by RNA-dependent RNA polymerases or by reverse transcription and PCR. This can take place using mRNAs separated from the cognate pairs, e.g., using poly T or poly U to hybridize to the poly A tails of, for instance, native unknown messages or by leaving the cognate pairs intact and using oligonucleotide primers that hybridize partially into the reading frame for known messages. Alternatively, commercial kits for rapid amplification of cDNA ends may be used. When this is used to evolve proteins and not just to select them, it would be preferable to sample at least one amino acid substitution at each position in the protein.

The Replication Threshold

A nominal minimum number of replications for efficient evolution may be estimated using the following formulae. If there is a sequence which is *n* sequences in length, with a selective improvement *r* mutations away with a mutation rate of *p*, the probability of generating the selective improvement on replication may be determined as follows:

For r =1, probability of a mutation at the right point, *p*, times the probability that it mutated to the right one of the three nucleotides that are different from the starting point, *1/3*, times the probability that the other n-1 sites remain unmutated, *(1-p)(n-r)*, or

$$P_r = \left(\frac{p}{3}\right)^1 (1-p)^{(n-1)}$$

where, P = the probability of attaining a given change r mutations away. More generally, for all *r* values:

$$P_r = \left(\frac{p}{3}\right)^r (1-p)^{(n-r)}$$

It is instructive to compare the chances of finding an advantage one mutation away with the chances three mutations away. This is because, given the triplet genetic code, any given codon can only change into nine other codons in one mutation. Indeed, it turns out that no codon can actually change into nine other amino acid codes in one mutation. The maximum number of amino acids that can be accessed in one mutation is seven amino acids and there are only eight codons of the sixty-four that can do this. Most codons have five or six out of nineteen other amino acids within one mutation. To reach all nineteen amino acids that are different from the starting one requires, in general, three mutations. These three mutations cannot be sequential since the two intervening ones will not, in general, be selectively advantageous. Therefore we need to use steps that are, at least, three mutations in size ($r=3$) to use all 20 amino acids.

For a mutation rate of .0067, which is that reported for "error-prone PCR", using a message of 300 nucleotides, which gives a short protein of 100 amino acids:

$$P_3 = 1.51 \times 10^{-9}$$

Therefore, one would expect to need a threshold of :

$$\frac{1}{1.51 \times 10^{-9}} = 6.64 \times 10^{8}$$

replications at that mutation rate to reasonably expect to reach the next amino acid that is advantageous. This is not the replication to use since the binomial expansion shows that over 1/3 of trials (actually about 1/e) would not contain the given sequence with selective advantage.

A poisson approximation for large n and small p for a given μ can be calculated so that we can compute the general term when n is, say, of the order 10^9 and p is of the order 10^{-9}. The general term of the approximation is:

$$\frac{\mu^r}{r!e^{\mu}}$$

An amplification factor of greater than approximately 6/P ensures that evolution will progress with the use of all amino acids. This is useful when the production of novel proteins precludes the use of "shuffling" of preexisting proteins.

Limits on Purification

Given a reversible binding where B and C compete for A:

$$AB \leftrightarrow A + B \qquad AC \leftrightarrow A + C$$

$$k_B = \frac{[A][B]}{[AB]} \qquad k_C = \frac{[A][C]}{[AC]}$$

$$[B] = k_B \frac{[AB]}{[A]} \qquad (1)$$

$$[C] = k_C \frac{[AC]}{[A]} \qquad (2)$$

The total concentrations can be expressed as follows:

$$[B]_T = [B] + [AB] \qquad (3)$$

$$[C]_T = [C] + [AC] \qquad (4)$$

Dividing (3) by (4):

$$\frac{[B]_T = [B] + [AB]}{[C]_T = [C] + [AC]}$$

And substituting (1) and (2) for [B] and [C]:

$$\frac{[B]_T = k_B \left[\frac{AB}{A}\right] + [AB]}{[C]_T = k_C \left[\frac{AC}{A}\right] + [AC]}$$

Rearranging the equation gives the following results:

$$\frac{[B]_T}{[C]_T} = \frac{[AB]\left(\frac{k_B + [A]}{[A]}\right)}{[AC]\left(\frac{k_C + [A]}{[A]}\right)}$$

Canceling the [A]'s in the numerator and denominator:

$$\frac{[B]_T}{[C]_T} = \frac{[AB](k_B + [A])}{[AC](k_C + [A])}$$

Finally, rearranging the equation provides the following equation:

$$\frac{[AB]}{[AC]} = \frac{[B]_T(k_C + [A])}{[C]_T(k_B + [A])}$$

$$\frac{(k_C + [A])}{(k_B + [A])} \qquad \text{(Enrichment Factor)}$$

The above factor is termed the "Enrichment Factor". The ratio of the total components is multiplied by this factor to calculate the ratio of the bound components, or the enrichment of B over C. The maximum enrichment factor is k_c/k_b, when the $[A]$ is significantly smaller than k_c or k_B.

The enrichment is limited by the ratio of binding constants. To enrich a scarce protein that is bound 100 times as strongly as its competitors, the ratio of that protein to its competitors is increased by 1 million with 3 enrichments. To enrich a protein that only binds twice as strongly than its competitors, 10 enrichment cycles would gain only an enrichment of ~1000.

By an exactly analogous method an enrichment factor of selecting proteins that bind least well can be shown:

In the equation:

$$\frac{[C]}{[B]} = \frac{k_C[C]_T([A] + k_B)}{k_B[B]_T([A] + k_C)}$$

The enrichment here is maximal at $[A]>k_A$ or k_B.

$$\frac{k_C\left([A]+k_B\right)}{k_B\left([A]+k_C\right)}$$

The following Examples illustrate various embodiments of the present invention and are not intended in any way to limit the invention.

5

EXAMPLE 1: PRODUCTION OF THE SATA USING URIDINE

One skilled in the art will understand that the SATA can be produced in a number of different ways. For example, in a preferred embodiment, three fragments (Fig. 1) were 10 purchased from a commercial source (i.e. Dharmacon Research Inc., Boulder, CO). Modified bases and a fragment 3 with a pre-attached puromycin on its 3' end and a PO_4 on its 3' end were included, all of which were available commercially. Three fragments were used to facilitate manipulation of the fragment 2 in forming the monoadduct.

Yeast tRNAAla or yeast tRNAPhe were used; however, sequences can be chosen 15 from widely known tRNA's or by selecting sequences that will form into a tRNA-like structure. Preferably, sequences with only a limited number of U's in the portion that corresponds to the fragment 2 are used. Using a sequence with only a few U's is not necessary because psoralen preferentially binds 5'UA3' sequences (Thompson J.F., et al Biochemistry 21:1363, herein incorporated by reference). However, there would be less 20 doubly adducted product to purify out if such a sequence was used.

Fragment 2 was preferably used in a helical conformation to induce the psoralen to intercalate. Accordingly, a complementary strand was required. RNA or DNA was used, and a sequence, such as poly C to one or both ends, was added to facilitate separation and removal after monoadduct formation was accomplished.

25 Fragment 2 and the cRNA were combined in buffered 50 mM NaCl solution. The Tm was measured by hyperchromicity changes. The two molecules were re-annealed and incubated for 1 hour with the selected psoralen at a temperature ~10°C less than the Tm. The psoralen was selected based upon the sequence used. A relatively insoluble psoralen, such as 8 MOP, could be selected which has a higher sequence stringency but may need to 30 be replenished. A more soluble psoralen, such as AMT, has less stringency but will fill

most sites. Preferably, HMT is used. If a fragment 2 is chosen that contains more non-target U's, a greater stringency is desired. Decreasing the temperature or increasing ionic strength by adding Mg^{++} was also used to increase the stringency. In a preferred embodiment, MG^{++} was omitted and ~400 mM NaCl solution was used.

Following incubation, psoralen was irradiated at a wavelength greater than approximately 400 nm. The irradiation depends on the wavelength chosen and the psoralen used. For instance, approximately 419 nm 20-150 J/cm2 was preferably used for HMT. This process results in an almost entirely furan sided monoadduct.

PURIFICATION OF A MONOADDUCT

The monoadduct was then purified by HPLC as described in Sastry et al, J. Photochem. Photobiol. B Biol. 14:65-79, herein incorporated by reference. The fact that fragment 2 was separate from fragment 3 facilitated the purification step because, generally, purification of monoadducts 25 mer is difficult (Spielmann et al. PNAS 89: 4514-4518, herein incorporated by reference).

LIGATION OF FRAGMENT 2 AND 3

The fragment 2 was ligated to the fragment 3 using T4 RNA ligase. The puromycin on the 3' end acted as a protecting group. This is done as per Romaniuk and Uhlenbeck, Methods in Enzymology 100:52-59 (1983), herein incorporated by reference. Joining of fragment 2+3 to the 3' end of fragment 1 was done according to the methods described in Uhlenbeck, Biochemistry 24:2705-2712 (1985), herein incorporated by reference. Fragment 2+3 was 5' phosphorylated by polynucleotide kinase and the two half molecules were annealed.

In an alternative method, significant quantities of furan sided monoadducted U were formed by hybridizing poly UA to itself and irradiating as above. The poly UA was then enzymatically digested to yield furan sided U which was protected and incorporated into a tRNA analog by nucleoside phosphoramidite methods. Other methods of forming the psoralen monoadducts include the methods described in Gamper et al., J. Mol. Biol. 197: 349 (1987); Gamper et al., Photochem. Photobiol. 40:29, 1984; Sastry et al, J. Photochem. Photobiol. B Biol. 14:65-79; Spielmann et al. PNAS 89:4514-4518, U.S. Patent Number 4,599,303, all herein incorporated by reference.

SATAs generated by the methods described above read UAG (anticodon CUA). Additionally, UAA or UGA were also used. In various embodiments, any message that had the stop codon that was selected as the "linking codon" was used.

PRODUCTION OF PSORALENATED FURAN SIDED MONOADDUCTS FROM CUAGA'PCUGGAGG RNA FRAGMENTS

UV Light Exposure Of RNA:DNA Hybrids

Equal volumes of 3 ng/ml RNA:cRNA hybrid segments and of 10 µg/ml HMT both comprised of 50mM NaCl were transferred into a new 1.5 ml capped polypropylene microcentrifuge tube and incubated at 37°C for 30 minutes in the dark. This was then transferred onto a new clean culture dish. This was positioned in a photochemical reactor (419 nm peak Southern New England Ultaviolet Co.) at a distance of about 12.5 cm so that irradiance was ~6.5 mW/cm2 and irradiated for 60-120 minutes.

Removal of Low Molecular Weight Protoproducts

100µl of chloroform-isoamyl alcohol (24:1) was pipetted and mixed by vortex. The mixture was centrifuged for 5 minutes at 15000 xg in a microcentrifuge tube. The chloroform-isoamyl alcohol layer was removed with a micropipet. The chloroform-isoamyl alcohol extraction was repeated once again. Clean RNA was precipitated out of the solution.

Alcohol Precipitation

Two volumes (~1000 µl) ice cold absolute ethanol was added to the mixture. The tube was centrifuged for 15 minutes at 15,000xg in a microcentrifuge. The supernatant was decanted and discarded and the precipitated RNA was redissolved in 100µl DEPC treated water then re-exposed to the RNA+8-MOP.

Isolation of the Psoralentated RNA Fragments Using HPLC

All components, glassware and reagents were prepared so that they were RNAase free. The HPLC was set up with a Dionex DNA PA-100 package column. The psoralenated RNA:DNA hybrid was warmed to 4°C. The psoralenated RNA was applied to HPLC followed by oligmucleotide analysis, as described in the following section entitled "Oligmucleotide Analysis by HPLC." The collected fractions represented:

- 5'CUAGAΨCUGGAGG3', where Ψ is pseudouridine (SEQ ID NO: 1)
- Furan sided 5'CUPsoralenAGAΨCUGGAGG3' monoadducts (SEQ ID NO: 2)
- 5'XXXXXCCUCCAGAUCUAGXXXXX3' (SEQ ID NO: 3)
- 5'XXXXXCCUCCAGAUCUPsoralenAGXXXXX3' (SEQ ID NO: 4)

The fractions were stored at 4°C in new, RNAase free snapped microcentrifuge tubes and stored at -20°C if more than four weeks of storage were required.

Identification of the RNA Fragments Represented by Each Peak Fraction Collected by HPLC Using Polyacrylamide Gel Electrophoresis (PAGE)

The electrophoresis unit was set up in a 4°C refrigerator. A gel was selected with a 2 mm spacer. Each 5 µl of HPLC fraction was diluted to 10 µl with Loading Buffer. 10 µl of each diluted fraction was loaded into appropriately labeled sample wells. The tracking dye was loaded in a separate lane and electrophoresis was run as described in the following section entitled "Polyacrylamide Gel Electrophoresis (PAGE) of Psoralenated RNA Fragments." After the electrophoresis run was complete, the electrophoresis was stopped when the tracking dye reached the edge of the gel. The apparatus was disassembled. The gel-glass panel unit was placed on the UV light box. UV lights were turned on. The RNA bands were identified. The bands appeared as denser shadows under UV lighting conditions.

Extraction of the RNA From the Gel

Each band was excised with a new sterile and RNAase free scalpel blade and transferred into a new 1.5 ml snap capped microcentrifuge tube. Each gel was crushed against the walls of the microcentrifuge tubes with the side of the scalpel blade. A new blade was used for each sample. 1.0 ml of 0.3M sodium acetate was added to each tube and eluted for at least 24 hours at 4°C. The eluate was transferred to a new 0.5 ml snap capped polypropylene microcentrifuge tube with a micropipet. A new RNAase free pipette tip was used for each tube and the RNA with ethanol was precipitated out.

Ethanol Precipitation

Two volumes of ice cold ethanol was added to each eluate then centrifuged at 15,000 xg for 15 minutes in a microcentrifuge. The supernatants were discharged and the precipitated RNA was re-dissolved in 100 µl of DEPC treated DI water. The RNA was stored in the microcentrifuge tubes at 4°C until needed. The tubes were stored at -20°C if storage was for more than two weeks. The following was order of rate of migration for each fragment in order from fastest to slowest:

a) 5'CUAGAΨCUGGAGG3' (SEQ ID NO:1)

b) Furan sided 5'CUPsoralenAGAΨCUGGAGG3' monoadducts (SEQ ID NO:2)

c) 5'XXXXXCCUCCAGAUCUAGXXXXX3' (SEQ ID NO: 3)

d) 5'XXXXXCCUCCAGAUCUPsoralenAGXXXXX3' (SEQ ID NO: 4)

The tubes containing the remainder of each fraction were labeled and stored at -20°C.

ETHANOL PRECIPITATION

RNA oligonucleotide fragments were precipitated, and all glassware was cleaned to remove any traces of RNase as described in the following section entitled "Inactivation of RNases on Equipment, Supplies, and in Solutions." All solutions were stored in RNAase free glassware and introduction of nucleases was prevented. Absolute ethanol was stored at 0°C until used. Micropipetors were used to add two volumes of ice cold ethanol to nucleic acids that were to be precipitated in microcentrifuge tubes. Capped microcentrifuge tubes were placed into the microfuge and spun at 15,000 xg for 15 minutes. The supernatant was discarded and precipitated RNA was re-dissolved in DEPC treated DI-water. RNA was stored at 4°C in microcentrifuge tubes until ready to use.

LIGATION OF RNA FRAGMENTS 2 AND 3

All glassware was cleaned to remove any traces of RNase as described in the following section entitled "Inactivation of RNases on Equipment, Supplies, and in Solutions." The following was added to a new 1.5 ml polypropylene snap capped

microcentrifuge tube using a 100-1000 µl pipet and a new sterile pipet tip was used for each solution:

Fragment 2 (3.0nM)	125.0µl
Fragment 3 (3.0nM)	125.0µl
Reaction buffer	250.0µl
RNA T4 ligase (9-12U/ml)	42µl

Reaction Buffer

RNase free DI-water	90.00ml
Tris-HCl (50mM)	0.79g
MgCl2 (10mM)	0.20g
DTT (5mM)	0.078g
ATP (1mM)	0.55g
pH to 7.8 with HCL	
RNase free DI-water	QS to 100.00ml

The mixture was gently mixed and the RNA was melted by incubating the mixture at 16°C for one hour in a temperature controlled refrigerated chamber. RNA was precipitated out of the solution immediately after the incubation was completed.

Alcohol Precipitation

Two volumes (~1000 µl) of ice cold absolute ethanol were added to the reaction mixture. The microcentrifuge tube was placed in a microcentrifuge at 15,000 xg for 15 minutes. The supernate was decanted and discarded and the precipitated RNA was re-dissolved in 100 µl DEPC treated water. The mixture was electrophoresed as described in the following section entitled "Polyacrylamide Gel Electrophoresis (PAGE) of Psoralenated RNA Fragments." The following was the order of rate of migration for each fragment in order from fastest to slowest:

a) Frag. 2

5'CUAGAΨCUGGAGG3'-OHPsoralen (SEQ ID NO: 5)

b) Frag. 3

5'UCCUGUGTΨCGAUCCACAGAAUUCGCACC-Puromycin (SEQ ID NO: 6)

c) Frag 2+3

5'CUPsoralenAGAYCUGGAGGUCCUGUGTΨCGAUCCACAGAAUUCGCACC Puromycin (SEQ ID NO: 7)

Each fraction was isolated by UV shadowing, the bands were cut out, the RNAs were eluted from the gels and the RNA elute was precipitated out as described in the following section entitled "Polyacrylamide Gel Electrophoresis (PAGE) of Psoralenated RNA Fragments." The ligation procedure was repeated with any residual unligated fragment 2 and 3 fractions. The ligated fractions 2 and 3 were pooled and stored in a small volume of RNase free DI-water at 4°C.

LIGATION OF RNA FRAGMENT 1 WITH FRAGMENT 2+3

All glassware was cleaned to remove any traces of RNase as described in the following section entitled "Inactivation of RNases on Equipment, Supplies, and in Solutions." The following was added to a new 1.5 ml polypropylene snap capped microcentrifuge tube. A 100-1000 µl pipet and new tip was used for each solution:

Fragment 2+3 (3.0nM)	125.0µl
Reaction buffer	250.0µl
T4 Polynucleotide Kinase(5-10U/ml)	1.7µl

Reaction Buffer

RNase free DI-water	90.00ml
Tris-HCl (40mM)	0.63g
MgCl2 (10mM)	0.20g
DTT (5mM)	0.08g
ATP (1mM)	0.006g
pH to 7.8 with HCL	
RNase free DI-water	QS to 100.00ml

The RNA was gently mixed then melted by heating the mixture to 70°C for 5 minutes in a heating block. The mixture was cooled to room temperature over a two hour period and the RNA was allowed to anneal in a tRNA configuration. The RNA was precipitated out of the solution.

Alcohol Precipitation

Two volumes (~1000 µl) of ice cold absolute ethanol were added to the reaction mixture. The microcentrifuge tube was placed in a microcentrifuge at 15,000 xg for 15 minutes. The supernatent was decanted and discarded and the precipitated RNA was re-dissolved in 100 µl DEPC treated water. The mixture was electrophoresed as described in the following section entitled "Polyacrylamide Gel Electrophoresis (PAGE) of Psoralenated RNA Fragments." The following was the order of rate of migration for each fragment in order from fastest to slowest:

a) Frag. 1

5'GCGGAUUUAGCUCAGDDGGGAGAGCGCCAGACU3' (SEQ ID NO: 8)

b) Frag 2+3

5'CUPsoralenAGAYCUGGAGGUCCUGUGTΨCGAUCCACAGAAUUCGCACC Puromycin (SEQ ID NO: 6)

c) Frag. 1+2+3

5'GCGGAUUUAGCUCAGDDGGGAGAGCGCCAGACUCUPsoralenAGAΨCU GGAGGUC...CUGUGTΨCGAUCCACAGAAUUCGCACCPuromycin (SEQ ID NO: 9)

Each fraction was isolated by UV shadowing, the bands were cut out, the RNAs were eluted from the gels and the RNA elute was precipitated out as described in the following section entitled "Polyacrylamide Gel Electrophoresis (PAGE) of Psoralenated RNA Fragments." The ligation procedure was repeated with the unligated Fragment 1 and

the 2+3 Fraction. The ligated fractions 2 +3 were pooled and stored in a small volume of RNase free DI-water at 4°C.

Final RNA Ligation

The following was added to a new 1.5ml polypropylene snap capped microcentrifuge tube. A 100-1000 µl pipet and new tip was used for each solution:

Fragment 1+2+3 (3.0nM)	250µl
reaction buffer	250µl
RNA T4 ligase (44µg/ml)	22µg

The mixture was incubated at 17°C in a temperature controlled refrigerator for 4.7 hours. Immediately after the incubation the tRNA was precipitated out as described in step 6.2 above and the tRNA was isolated by electrophoresis as described in the following section entitled "Polyacrylamide Gel Electrophoresis (PAGE) of Psoralenated RNA Fragments." The tRNA was pooled in a small volume of RNase free water and stored at 4°C for up to two weeks or stored at -20°C for periods longer than two weeks.

POLYACRYLAMIDE GEL ELECTROPHORESIS (PAGE) OF PSORALENATED RNA FRAGMENTS

Acrylamide Gel Preparation

All reagents and glassware were made RNAase free as described in the following section entitled "Inactivation of RNases on Equipment, Supplies, and in Solutions." The gel apparatus was assembled to produce a 4 mm thick by 20 cm x 42 cm square gel. 29 parts acrylamide with 1 part ammonium crosslinker were mixed at room temperature with the appropriate amount of acrylamide solution in an RNAase free, thick walled Erlenmeyer flask.

Acrylamide Solution

urea (7M)	420.42 g
TBE (1X)	QS to 1L

5X TBE

0.455 M Tris-HCl	53.9g
10mM EDTA	20ml of 0.5 M
RNAase free DI water	900ml
pH with boric acid to	pH 9
QS with RNAase free DI water to	1L

The mixture was degassed with vacuum pressure for one minute. The appropriate amount of TEMED was added, mixed gently, and then the gel mixture was poured between the glass plates to within 0.5 cm of the top. The comb was immediately inserted between the glass sheets and into the gel mixture. An RNAase free gel comb was used. The comb produced wells for a 5 mm wide dye lane and 135 mm sample lanes. The gel was allowed to polymerize for about 30-40 minutes then the comb was carefully removed. The sample wells were rinsed out with a running buffer using a micropipet with a new pipet tip. The wells were then filled with running buffer.

Sample Preparation

An aliquot of the sample was suspended in loading buffer in a snap capped microcentrifuge tube and vortex mixed. Indicator dye was not added to the sample.

Loading Buffer	
Urea (7M)	420.42 g
Tris HCl (50mM)	7.85 g
QS with RNAase free D-H2O	to 1L

Electrophoresis run

The maximum volume of RNA/loading buffer solution was loaded into the 135 mm sample wells and the appropriate volume of tracking dye in 5 mm tracking lane. The samples were electrophoresed in a 5°C refrigerator. The electrophoresis was stopped when the tracking dye reached the edge of the gel. The apparatus was then disassembled. Glass panels were not removed from the gel. The gel-glass panel unit was placed on a UV light box. With UV filtering goggles in place, the UV lights were turned on. The RNA bands were identified. They appeared as denser shadows under UV lighting conditions. The RNA was extracted from the gel. Each band was excised with a new sterile and RNAase

free scalpel blade and each band was transferred into a new 1.5 ml snap capped microcentrifuge tube. Each gel was crushed against the walls of the microcentrifuge tubes with the side of the scalpel blade. A new blade was used for each sample. 1.0 ml of 0.3M sodium acetate was added to each tube and eluted for at least 24 hours at 4°C. The eluate was transferred to a new 0.5 ml snap capped polypropylene microcentrifuge tubes with a micropipet with a new RNAase free pipet tip for each tube. Two volumes of ice cold ethanol was added to each eluate, then centrifuged at 15,000 xg for 15 minutes in a microcentrifuge. The supernatants were discarded and the precipitated RNA was redissolved in 100 µl of DEPC treated DI water. The RNA was stored in the microcentrifuge tubes at 4°C until needed.

OLIGONUCLEOTIDE ANALYSIS BY HPLC

HPLC purification of the RNA oligonucleotides was performed using anion exchange chromatography. Either the 2'-protected or 2'-deprotected forms may be chromatographed. The 2'-protected form offered the advantage of minimizing secondary structure effects and providing resistance to nucleases. If the RNA was fully deprotected, sterile conditions were required during purification.

One skilled in the art will understand that the HPLC purification methods of Example 2 may be modified in order to purify the RNA oligonucleotides. Modification of the HPLC purification methods of Example 2, including HPLC gradient, temperature, and other parameters, may be necessary. One of skill in the art would also recognize that a one-step HPLC purification method may also be used in accordance with several embodiments of the current invention.

INACTIVATION OF RNASES ON EQUIPMENT, SUPPLIES, AND IN SOLUTIONS

Glassware was treated by baking at 180°C for at least 8 hours. Plasticware was treated by rinsing with chloroform. Alternatively, all items were soaked in 0.1% DEPC.

Treatment with 0.1% DEPC

0.1% DEPC was prepared. DI water was filtered through a 0.2µM membrane filter. The water was autoclaved at 15 psi for 15 minutes on a liquid cycle. 1.0g (wt/v) DEPC/liter of sterile filtered water was added.

Glass and Plasticware

All glass and plasticware was submerged in 0.1% DEPC for two hours at 37°C. The glassware was rinsed at least 5X with sterile DI water. The glassware was heated to 100°C for 15 minutes or autoclaved for 15 minutes at 15 psi on a liquid cycle.

Electrophoresis Tanks Used for Electrophoresis of RNA

Tanks were washed with detergent, rinsed with water then ethanol and air dried. The tank was filled with 3% (v/v) hydrogen peroxide (30ml/L) and left standing for 10 minutes at room temperature. The tank was rinsed at least 5 times with DEPC treated water.

Solutions

All solutions were made using Rnase free glassware, plastic ware, autoclaved water, chemicals reserved for work with RNA and RNase free spatulas. Disposable gloves were used. When possible, the solutions were treated with 0.1% DEPC for at least 12 hours at 37°C and then heated to 100°C for 15 minutes or autoclaved for 15 minutes at 15 psi on a liquid cycle.

RNA TRANSLATION

2 µl of gastroinhibitory peptide (GIP) mRNA at a concentration of 20 µl/ml was placed in a 250 µl snapcap polypropylene microcentrifuge tube. 35 µl of rabbit reticulocyte lysate (available commercially from Promega) was added. 1 µl of amino acid mixture which did not contain methionine (available commercially from Promega) was added. 1 µl of 35S methione or unlabeled methionine was added. Optionally, 2 ml of luciferase may be added to some tubes to serve as a control. In a preferred embodiment, luciferase was used instead of GIP mRNA. One skilled in the art will understand that indeed any mRNA fragment containing the appropriate sequences may be used.

SATA was added to the experimental tubes. Control tubes which did not contain SATA were also prepared. The quantity of SATA used was approximately between 0.1µg to 500 µg, preferably between 0.5 µg to 50 µg. 1 µl of Rnasin at 40 units/ml was added. Nuclease free water was added to make a total volume of 50 µl.

For proteins greater than approximately 150 amino acids, the amount of tRNA may need to be supplemented. For example, approximately 10 - 200 μg of tRNA may be added. In general, the quantity of the SATA should be high enough to effectively suppress stop or pseudo stop codons. The quantity of the native tRNA must be high enough to out compete the SATA which does not undergo dynamic proofreading under the action of elongation factors.

Each tube was immediately capped, parafilmed and incubated for the translation reactions at 30°C for 90 minutes. The contents of each reaction tube was transferred into a 50 μl quartz capillary tube by capillary action. The SATA was crosslinked with mRNA by illuminating the contents of each tube with 2-10 J/cm2 ~350 nm wavelength light, as per Gasparro et al. (Photochem. Photobiol. 57:1007 (1993), herein incorporated by reference). Following photocrosslinking, the contents of each tube were transferred into a new snapcap microfuge tube. The ribosomes were dissociated by chelating the calcium cations by adding 2 μl of 10 mM EDTA to each tube. Between each step. each tube was gently mixed by stirring each component with a pipette tip upon addition.

The optimal RNA for a translation was determined prior to performing definitive experiments. Serial dilutions may be required to find the optimal concentration of mRNA between 5-20 μg/ml.

Reagent	1	2	3	4
Rabbit reticulocyte lysate (35µl)	+	+	+	+
Amino acid mixture minus methionine (1µl of 1 mM)	+	+	+	+
^{35}S Methionine (1µl of 1,200Ci/mmol)	+	-	-	+
Methionine (unlabeled)	-	+	+	-
GlP mRNA (2µl of 20 µg/ml)	+	-	-	-
^{32}P GlP mRNA (2µl of 20 µg/ml)	-	+	+	-
Rnasin (1µl of 40 U/µl)	+	+	+	+
SATA	-	-	-	
Water, nuclease free (q.s. to 50 µl)	+	+	+	+

SDS-Page electrophoresis was performed on each sample, as described above. Autoradiography on the gel was performed, as described by Sambrook et. al., Molecular Cloning, A Laboratory Manual, 2nd ed., Coldspring Harbor Press (1989), herein incorporated by reference.

In another example, the SATA was produced in a manner similar to the above methodology, except that uridines were substituted with pseudouridines. This technique is discussed below in Example 2.

EXAMPLE 2: PRODUCTION OF THE SATA USING PSEUDOURIDINE

As discussed above, one skilled in the art will appreciate that the SATA can be produced in a number of different ways. Figure 5 shows the chemical structures for uridine and pseudouridine. Pseudouridine is a naturally occurring base found in tRNA that forms hydrogen bonds just as uridine does, but lacks the 5-6 double bond that is the target for psoralen. In a preferred embodiment, the SATA was produced using pseudouridine. Specifically, in a preferred embodiment, three fragments (Fig. 1) were purchased from a commercial source (Dharmacon Research Inc., Boulder, CO). Modified bases and a fragment 3 ("Fragment 3") with a pre-attached puromycin on its 3' end and a PO_4 on its 3'

end were included, all of which are available commercially. The three fragments were used to facilitate manipulation of a fragment 2 ("Fragment 2") in forming the monoadduct. Preferred sequences of the three fragments are as follows:

Fragment 1

5'PO_4GCGGAUUUAGCUCAGUUGGGAGAGCGCCAGACOH3' (SEQ ID NO: 10)

Fragment 2

5'OHΨCUAACΨCOH3' (SEQ ID NO: 11)

Fragment 3

5'PO_4UGGAGGUCCUGUGdTΨCGAUCCACAGAAUUCGCACCPuromycin3' (SEQ ID NO: 12)

Yeast tRNAAla or yeast tRNAPhe was preferably used. However, one skilled in the art will understand that sequences can be chosen widely from known tRNAs or by selecting sequences that will form into a tRNA-like structure. Using pseudouridine instead of uridine in Fragment 2 avoids psoralen labeling of the nontarget U's.

Fragment 2 was preferably used in a helical conformation to induce the psoralen to intercalate. Accordingly, a complementary strand was required. RNA or DNA was used, and a sequence, such as poly C to one or both ends, was added to facilitate separation and removal after monoadduct formation was accomplished. Use of pseudouridine instead of uridines in the complement means that a high efficiency wave length, such as 365 nm, can be used with out fear of crosslinking the product. Irradiation was preferably in the range of about 300-450 nm, more preferably in the range of about 320 to 400 nm, and most preferably about 365 nm. Further, use of pseudouridine left the furan sided monoadduct in place on Fragment 2 because the Maf is the predominate first step in the crosslink formation.

The following cRNA sequence with pseudo uridine was used according to a preferred embodiment of the present invention. :

CCCΨCCAGAGΨΨAGACCC (SEQ ID NO: 13)

Step 1: Furan sided monoadduction of psoralen to fragment 2

The formation of a furan sided psoralen monoadduct with the target uridine of Fragment 2 was performed as follows:

A reaction buffer was prepared as follows:

Tris HCL	25 mM
NaCl	100 mM
EDTA	0.32 mM
pH	7.0

4'hydroxy methyl-4,5',8'-triethyl psoralen (HMT) was then added to a final concentration of 0.32 mM and equimolar amounts of fragment 2 and cRNA were added to a final molar ratio of fragment 2: cRNA : psoralen = 1:1:1000. A total volume of 100µl was irradiated at a time.

The mixture of complementary oligos, HMT, psoralen was processed as follows:

1) Heated to 85° C for 60 sec followed by cooling to 4° C over 15 min, using PCR thermocycler.

2) Irradiated for 20 min at 4° C, in Eppendorf UVette plastic cuvette, covered top with parafilm, laid on the top of UV lamp (1mW/cm^2 multi-wavelength UV lamp (λ>300nm) (UV L21 model λ 365 nm).

Steps 1 and 2 above were repeated 4 times to re-intercalate and irradiate HMT. After the second irradiation additional 10 µl of 1.6 mM HMT was added in total 100µl reaction volume. After 4 cycles of irradiation, the free psoralens were extracted with chloroform and all oligos (labeled and unlabeled) were precipitated with ethanol overnight (see precipitation step). A small aliquot was saved for gel identification.

Step 2: Purification of HMT conjugated fragment 2 (2MA) oligo by reverse-phase HPLC

1) The reaction mixture was dried with speed vacuum for 10 minutes and then was dissolved with 2 µl of 0.1 M TEAA, pH 7.0 buffer.

0.1 M TEAA, pH 7.0 Buffer

Acetic Acid	5.6 ml
Triethylamine	13.86 ml
$H_2$0 (RNAase free)	950 ml

pH adjusted to 7.0 with acetic acid

and water added to 1L

2) The sample was loaded onto a Waters Xterra MS C18, 2.5 μm, 4.5x50 mm reverse-phase column pre-equilibrated with buffer A (5% wt/wt acetonitrile in 0.1M TEAA, pH 7.0) The sample was eluted with a gradient of 0-55% buffer B (15% wt/wt acetonitrile in 0.1M TEAA, pH 7.0) to buffer A over a 35 minute time frame at a flow rate of 1 ml/minute. The column temperature was 60°C and the detection wave length, set by a narrow band filter, was 340 nm. Furan sided psoralen monoadduct absorbs at 340 nm but the RNA, and any pyrone sided monoadduct does not. The buffer solutions were filtered and degassed before use.

The 2MA eluted at around 25-28 minutes at a buffer B concentration of 40%. Unpsoralenated fragment 2 eluted before 8 minutes based on subsequent gel electrophoresis analysis on collected fractions.

The column was washed with 100% acetonitrile for 5 minutes and was re-equilibrated with buffer A for 15 minutes. All fractions were dried with speed vacuum overnight.

The fractions containing the 2MA were identified by the level of absorbance at 260 nm (RNA) and 330 nm (furan sided psoralen monoadducted RNA). This was done by redissolving the dried fractions with 120 μl of Rnase-free distilled water and the absorbance was measured with a spectrophotometer at 260 nm and 330 nm. The fractions with high absorbance at both wavelengths were pooled then dried with speed vacuum. A small aliquot from each was saved for gel analysis.

The cross-linked products were analyzed on a denaturing 20% TBE-urea gel and visualized by gel silver staining.

Step 3: Purification of HMT conjugated fragment 2 oligo from cRNA by anion exchange HPLC

The dried samples were pooled and then were dissolved with 0.5X TE buffer. A sample of about 0.4 absorbance unit was loaded onto a Dionex DNAPac PA-100 (4x250mm) column which was pre-equilibrated with buffer C (25 mM Tris-HCl, pH 8.0) and the column temperature was 85°C.

The oligos were eluted at a flow rate of 1 ml/min. with a concave gradient from 4% to 55% buffer D for 15 minutes followed by a convex gradient from 55 % to 80% with

buffer D for the next 15 minutes. The oligos were washed with 100% buffer D for 5 min and 100% buffer C for another 5 min at a flow rate of 1.5 ml/min; Fractions were collected that absorbed 260 nm light. 2MA had a retention time (RT) of 16.2 minutes and was eluted by 57% buffer D, and free fragment 2 had RT less than 16.6 minutes, and was eluted by 55% buffer D and free cRNA had RT greater than 19.2 minutes. The fractions were collected that absorbed at 254 or 260 nm. The collected fractions were dried with speed vacuum overnight. All solutions were filtered and degassed before use.

The solution used comprised the following:

C: 25mM Tris-HCl pH 8.0;

D: 250mM NaClO4 in 25mM Tris pH 8.0 buffer.

TE: 10mM Tris-HCl pH 8.0 with 1mM EDTA

Step 4: Desalting, precipitation and collection of the purified 2MA oligo

The dried fractions were redesolved with 100µl Rnase free distilled water. 500µl cool 100% ethanol with 0.5M (NH4)2CO3 was added and the mixture was vortexed briefly. The mixture was then frozen on dry ice for 60 minutes or stored at -20°C overnight.

The samples were then brought to 4° C and centrifuged at maximum speed in a microcentrifuge for 15 minutes. The position of the pellet was noted and the supernatant was decanted or removed by pipette. Care was taken not to disturb pellet. If the pellet still contained salt, this step was repeated. The pellet was then washed with 70% pre-cooled ethanol twice. The wet pellet was dried with speed vacuum for 15 min.

Urea PAGE gel identified the right fractions for the next step.

Step 5: Ligation of 2MA oligo to Fragment 3 oligo:

A. The following reagents and instruments were used:

- Nuclease-Free Water (Promega)
- polyethylene glycol (PEG8000 Sigma) 40%(wt/wt in water)
- RNasin ® Ribonuclease Inhibitor (Promega)
- phenol:chloroform
- 1.5 ml sterile microcentrifuge tubes
- 100% ethanol

- 70% ethanol
- Dry ice or –20°C freezer
- Microcentrifuge at room temperature and +4°C
- PCR thermocycler or water bath

B. The following reaction conditions were used:

- 50 mM Tris-HCl (pH 7.8),
- 10 mM MgCl2,
- 10 mM DTT
- 1 mM ATP and
- 18-20% PEG

C. The following reaction mixture was assembled in a sterile microcentifuge tube:

- Fragment 3 (Donor) 1µl (6µg) (Purified, when necessary, before using as a donor)
- 2MA (Acceptor) 1µl (1.5µg

After adding 8 µl Rnase free dH2O 8µl, the reactions were incubated at 85° C for 1 minute to relax the oligo secondary structure, then slowly cooled to 4° C, using a PCR machine thermocycler. The preheated tube was placed on ice to keep cool and centrifuged briefly, then the following was added:

- 10X Ligase Buffer 4µl
- 10mM ATP 4µl
- Rnase Out or Rnasin(40u/µl) Promega 0.5µl
- PEG, 40 % (Sigma) 20µl
- T4 RNA Ligase (10u/µl) (NEB) 1µl

Nuclease-free water was added to final Volume of 40 µl. The mixture was incubate at 16° C overnight(16hr). The mixture was centrifuged briefly and then was placed on ice.

D. Precipitation of Oligonucleotides:

60µl DEPC RNase free distilled water was added to the mixture and then 150 µl phenol/chloroform was added. The mixture was vortexed vigorously for 30 seconds. The precipitate was then centrifuged out at maximum speed in a microcentifuge for 5 minutes at

room temperature. The aqueous phase was transferred to a new microcentrifuge tube (>95µl).

To this was added 3µl 5 mg/ml glycogen, and 500µl pre-cooled 100% ethanol with 0.5M (NH4)2CO3 and the mixture was vortexed briefly and then was frozen on dry ice for 60 minutes. At this point, it may be stored overnight at -20°C. The dried fractions were redissolved with 100µl Rnase-free distilled water, 500µl cool 100% ethanol with 0.5M (NH4)2CO3 was added and vortexed briefly. This was then frozen on dry ice for 60 minutes or stored at -20C overnight. The samples were then brought to 4° C and centrifuged at maximum speed in a microcentrifuge for 15 minutes and supernatant removed by pipette. Care was taken not to disturb pellet. If the pellet still contained salt, this step was repeated once. The pellet was then washed with 70% pre-cooled ethanol several times. This was then centrifuged at maximum speed in a microcentrifuge for 5 minutes at 4C. The ethanol was carefully removed using a pipette. Centrifugation was repeated again to collect remaining ethanol which was carefully removed. The wet pellet was dried with speed vacuum for 10 min. A small aliquot was collected for the gel analysis. For long term storage, the RNA was stored in ethanol at -20C. Care was taken not to store the RNA in DEPC water.

Step 6: Purification of the ligated fragment 3 oligo complex

The dried sample was redesolved with 0.5X TE buffer and was loaded onto a DNAPac PA-100 column which was equilibrated with buffer C. The column temperature was 85°C and the detector operated at 254 nm to identify fractions with RNA and at 340 nm to identify fractions with 2MaF. The oligos were eluted with a convex gradient from 30% to 70% with buffer D for the first 20 minutes at a flow rate of 0.8ml/min and followed with a linear gradient from 70 % to 98% D for another 20 min at the same flow rate. The elution was completed by washing with 100% D for 7 min and 100% C for another 10 min at 1.0 ml/min flow rate. The fractions were detected with 254 or 260 nm wavelength light. The ligated oligos (2MA-fragment 3) were eluted after 34 min, by more than 90% buffer B. Fractions with 254 nm absorbance (A_{254nm}>0.01) were collected and dried with speed vacuum overnight.

Step 7: Purified 2MA-fragment 3 desalting and precipitation:

The dried fractions were re-dissolved with 100µl Rnase free distilled water, 500µl cool 100% ethanol with 0.5M (NH4)2CO3 was added and the mixture was vortexed briefly. The mixture was then frozen on dry ice for 60 minutes or stored at -20C overnight.

The samples were brought to 4° C and centrifuged at maximum speed in a microcentrifuge for 15 minutes. The position of the pellet was noted and the supernatant decanted or removed by pipette. Care was taken not to disturb pellet. If still containing salt, this step was repeated. The pellet was then washed with 70% pre-cooled ethanol twice. The wet pellet was dried with speed vacuum for 15 min.

Urea PAGE was performed to identify the ligated 2MA-fragment-3 for use in the next step of ligating fragment 1 to the 2MA-fragment-3 oligo which completes the SATA linker.

Step 8: Preparation of SATA

A. RNA Oligo 5'phosphorylation

1. Reagent and instrument:

- Nuclease-Free Water (Cat.# P1193 Promega)
- RNasin ® Ribonuclease Inhibitor (Cat# N2511 Promega)
- Phenol:chloroform
- Sterile microcentrifuge tubes
- 100% ethanol
- 70% ethanol
- Microcentrifuge at room temperature and +4°C
- PCR thermalcycler or water bath

2. Assemble the following reaction mixture in a sterile microcentrifuge tube:

Component	Volume
• Acceptor RNA	<200ng
• T4 ligase 10X Reaction Buffer*	4µl
• RNasin ® Ribonuclease Inhibitor (40u/µl)	20unit
• T4 kinase (9-12u/µl)	2µl
• 10 mM ATP	4µl
• Nuclease-Free Water to final volume	40µl

Incubate at 37°C for 30 minutes in a PCR thermocycler or water bath. For non-radioactive phosphorylation, use up to 300 pmol of 5' termini in a 30 to 40 µl reaction containing 1X T4 Polynucleotide Kinase Reaction Buffer, 1 mM ATP and 10 to 20 units of T4 Polynucleotide Kinase. Incubate at 37°C for 30 minutes. 1X T4 DNA Ligase Reaction Buffer contains 1 mM ATP and can be substituted in non-radioactive phosphorylations. T4 Polynucleotide Kinase exhibits 100% activity in this buffer). Fresh buffer is required for optimal activity (in older buffers, loss of DTT due to oxidation lowers activity.

B. Annealing Fragment1 and 2MA-fragment 3 oligo complex:

1. Reagents and instruments:

- PCR thermocycler instrument or water bath
- 100µg/ml nuclease-free albumin
- 100mM MgCl2

2. Assemble the following reaction mixture in a sterile microcentrifuge tube:

- Acceptor RNA oligo (1E) <200 ng
- Donor RNA oligo (3G-2G ligated oligo) <200 ng
- (5' phosphorylated oligo from step A)

Appropriate ratios are as follows: Acceptor oligo:Donor oligo (Fragment 1: 2MA-Fragment 3) molar ratio should be 1:1.1 to avoid fragment 1 self-ligation. $MgCl_2$ was added to T4 ligase buffer (50mM Tris-HCl, (pH 7.8), 10 mM $MgCl_2$, 10 mM DTT and 1mM ATP) to final 20 mM concentration. Add Rnase free albumin to final 5µg/ml. The final volume should be no more than 100µl. The solution was heated to 70° C for 5 min, then was cooled from 70° C to 26° C over 2 hours and cooled from 26°C to 0°C over 40 minutes. Incubate at 16°C for 16 to 17 hours using PCR instrument.

C. Ligation of annealed oligos:

- Annealed oligos < 15µl
- 10mM ATP 2µl
- 40% PEG 18µl
- T4 ligase 10X Buffer 2µl

- RNasin ® Ribonuclease Inhibitor (40u/µl) 0.5µl
- T4 ligase (9-12u/µl)(NEB) 1µl
- Nuclease-Free Water to final volume 40µl

D. Precipitating tRNA fragment:

After ligation, 50 µl DEPC water and 150 µl phenol: chloroform were added and vortexed vigorously for 30 seconds. This was then centrifuged at maximum speed in a microcentrifuge for 5 minutes at room temperature. The aqueous phase was transferred to a new microcentrifuge tube (~100 µl). To this was added 2 µl 10mg/ml mussel glycogen, 10µl 3M sodium acetate, pH 5.2. This was mixed well. Then 220µl 95% ethanol was added and vortexed briefly. The mixture was then frozen on dry ice for 30 minutes. At this point the mixture may be stored over night at –20°C or one may proceed. The RNA must not be stored in DEPC water, but in ethanol, at –20° C.

Then the samples were brought to 4°C and centrifuged at maximum speed in a microcentrifuge for 15 minutes. The position of the pellet was noted and the supernatant decanted or removed by pipette. Care was taken not to disturb pellet. The pellet was then washed with 70% pre-cooled ethanol twice. After removing the ethanol, the wet pellet was dried with a speed vacuum for 15 min. The dried pellet was stored at –20°C, until the next step.

RNA Translation

A luciferase mRNA which was modified to have the stop codon corresponding to that recognized by the anticodon of the SATA (in the present case UAG) was used in a standard Promega *in vitro* translation kit in the recommended 1µl of concentration 1µg/µl. One skilled in the art will understand that indeed any mRNA fragment containing the appropriate sequences may be used.

SATA was added to the experimental tubes. Control tubes which did not contain SATA were also prepared. The quantity of SATA used was approximately between 0.1µg to 500 µg, preferably between 0.5 µg to 50 µg. 1 µl of Rnasin at 40 units/ml was added. Nuclease free water was added to make a total volume of 50 µl.

For proteins greater than approximately 150 amino acids, the amount of tRNA may need to be supplemented. For example, approximately 10 - 200 µg of tRNA may be

added. In general, the quantity of the SATA should be high enough to effectively suppress stop or pseudo stop codons. The quantity of the native tRNA must be high enough to out compete the SATA which does not undergo dynamic proofreading under the action of elongation factors.

Each tube was immediately capped, parafilmed and incubated for the translation reactions at 30°C for 90 minutes. The contents of each reaction tube was transferred into a 50 µl quartz capillary tube by capillary action. The SATA was crosslinked with mRNA by illuminating the contents of each tube with 2-10 J/cm2 ~350 nm wavelength light, as per Gasparro et al. (Photochem. Photobiol. 57:1007 (1993), herein incorporated by reference). Following photocrosslinking, the contents of each tube were transferred into a new snapcap microfuge tube. The ribosomes were dissociated by chelating the calcium cations by adding 2 µl of 10 mM EDTA to each tube. Between each step, each tube was gently mixed by stirring each component with a pipette tip upon addition.

The optimal RNA for a translation was determined prior to performing definitive experiments. Serial dilutions may be required to find the optimal concentration of mRNA between 5-20 µg/ml.

SDS-Page electrophoresis was performed on each sample, as described above. Autoradiography on the gel was performed, as described by Sambrook et al., Molecular Cloning, A Laboratory Manual, 2nd ed., Coldspring Harbor Press (1989), herein incorporated by reference.

EXAMPLE 3: ALTERNATIVE SEQUENCES

In a preferred embodiment, Fragments 1, 2 and 3, described above in Example 1, have the following alternate sequences:

Fragment 1 (SEQ ID NO: 13):

5' PO4 GCGGAUUUAGCUCAGUUGGGAGAGCGCCAGA N3-Methyl-U 3'

Fragment 2 (SEQ ID NO: 14):

5' UCUAAGΨCΨGGAGG 3'

Fragment 3

(Unchanged from the sequence listed above (SEQ ID NO: 6)):

5' PO4 UCCUGUGTΨCGAUCCACAGAAUUCGCACC Puromycin 3'

Using the methods described above, the sequence of alternative Fragments 1+2+3 was (SEQ ID NO: 15):

5'PO4 GCGGAUUUAGCUCAGUUGGGAGAGCGCCAGA N3-Methyl UUCUPsoralenAAGΨCΨGGAGGUCCUGUGTYCGAUCCACAGAAUUCG Puromycin 3'

While a number of preferred embodiments of the current invention and variations thereof have been described in detail, other modifications and methods of use will be readily apparent to those of skill in the art. Accordingly, it should be understood that various applications, modifications and substitutions may be made without departing from the spirit of the invention or the scope of the claims.

WHAT IS CLAIMED IS:

1. A method for selecting a desired protein or nucleic acid molecule comprising:

providing at least two candidate mRNA molecules, wherein at least one of said mRNA molecules contains at least one codon selected from the group consisting of a stop codon and a pseudo stop codon;

translating at least two of said candidate mRNA molecules to generate at least one translated protein;

linking at least one of said candidate mRNA molecules to its corresponding translated protein to form at least one cognate pair;

selecting one or more of said cognate pairs based upon the properties of said translated protein or said mRNA molecule; and

selecting said translated protein or said desired nucleic acid molecule comprising identifying a molecule selected from the group consisting of an mRNA molecule of said selected cognate pair, a nucleic acid molecule complementary to said mRNA molecule and a nucleic acid molecule homologous to said mRNA molecule.

2. The method of Claim 1, wherein at least one of said candidate mRNA molecules and at least one said translated protein is linked by a tRNA molecule selected from the group consisting of tRNA, modified tRNA and tRNA analogs.

3. The method of Claim 2, wherein said tRNA molecule is connected to said translated protein by ribosomal peptidyl transferase.

4. The method of Claim 2, wherein said tRNA molecule is connected to the mRNA through an ultraviolet-induced cross link between the anticodon of said tRNA molecule and the codon of the RNA message.

5. The method of Claim 2, wherein said tRNA molecule is connected to the mRNA by thiouracil.

6. The method of Claim 2, wherein said tRNA molecule is connected to the mRNA by a psoralen cross link.

7. The method of Claim 2, wherein said tRNA molecule is connected to the mRNA by using a psoralen monoadduct.

8. The method of Claim 7, wherein said psoralen monoadduct is placed on said mRNA.

9. The method of Claim 7, wherein said psoralen monoadduct is placed on an anticodon of said tRNA molecule.

10. The method of Claim 7, wherein said psoralen monoadduct is placed on a stop anticodon of said tRNA molecule.

11. The method of Claim 7, wherein said psoralen monoadduct is generated by forming crosslinks and photo reversing said crosslinks or by using visible light.

12. The method of Claim 7, wherein said psoralen monoadduct is generated using light of a wavelength in the range of about 300 to about 450 nm.

13. The method of Claim 12, wherein said wavelength is 365 nm.

14. The method of Claim 2, wherein said tRNA molecule has a stable peptide acceptor.

15. The method of Claim 2, wherein said tRNA molecule is modified to accept a peptide chain and hold said chain in a stable manner, such that peptidyl transferase cannot detach it.

16. The method of Claim 2, wherein said tRNA molecule is modified by using a bond which binds to the ribosome, accepts the peptide chain, and then does not act as a donor in the next transpeptidation.

17. The method of Claim 16, wherein said bond is selected from the group consisting of a 2' ester on a 3' deoxy adenosine and an amino acyl $tRNA_{ox\text{-}red}$.

18. The method of Claim 2, wherein an amino acid or amino acid analog is attached to the 3' end of said tRNA molecule by a stable bond to generate a stable aminoacyl tRNA analog.

19. The method of Claim 1, wherein said translating is performed *in vitro*.

20. The method of Claim 1, wherein said translating is performed in situ.

21. The method of Claim 1, wherein said cognate pair is selected based upon ligand binding.

22. The method of Claim 1, wherein selecting said desired nucleic acid molecule comprises:

having an array of nucleic acids, wherein said nucleic acids are placed in a predetermined position;

hybridizing at least one of said cognate pairs onto said array; and

identifying the mRNA based upon its reaction to a protein identifier.

23. The method of Claim 1, wherein said reaction to a protein identifier is selected from the group consisting of ligand binding, immunoprecipitation and enzymatic reactions.

24. A kit to generate cognate pairs comprising at least one psoralen monoadduct attached to a nonadducted stable aminoacyl tRNA analog or at least one psoralen monoadduct attached to an oligonucleotide.

25. A method for evolving a desired protein sequence comprising:

providing at least two candidate mRNA molecules, wherein at least one of said mRNA molecules contains at least one codon selected from the group consisting of a stop codon and a pseudo stop codon;

translating at least two of said candidate mRNA molecules to generate at least one translated first protein;

linking at least one of said candidate mRNA molecules to its corresponding translated first protein to form at least one first cognate pair;

selecting at least one of said first cognate pairs based upon at least one desired characteristic of said translated first protein or said mRNA;

recovering at least one of said first cognate pairs with said desired characteristic to generate at least one recovered cognate pair;

amplifying a first nucleic acid component of one or more of said recovered cognate pairs;

producing at least one second nucleic acid component comprising at least one of said first nucleic acid components with one or more variations;

producing at least one second protein by translating at least one of said second nucleic acid components;

linking at least one of said second proteins with at least one of said second nucleic acid components to generate one or more second cognate pairs; and

obtaining the desired protein sequence by re-selecting one or more of said second cognate pairs based upon at least one desired property, wherein said desired property is the same or different than said desired characteristic.

26. The method of Claim 25, wherein said desired characteristic is selected from the group consisting of binding properties, enzymatic reactions and chemical modifications.

27. The method of Claim 25, wherein selecting at least one of said first cognate pairs comprises:

providing a first ligand with a desired binding characteristic;

contacting one or more of said first cognate pairs with said first ligand to generate unbound complexes and bound complexes;

recovering either the bound complexes or the unbound complexes;

amplifying at least one nucleic acid component of the recovered complexes;

introducing variation to a sequence of said nucleic acid component of said recovered complexes;

translating one or more second proteins from said nucleic acid components,

linking at least one of said second proteins with at least one of said second nucleic acid components to generate one or more second cognate pairs; and

obtaining the desired protein sequence by contacting said at least one of said second cognate pairs with at least one second ligand to select one or more of said second cognate pairs, wherein said second ligand is the same or different than said first ligand.

28. A method of forming a monoadduct comprising:

providing a target oligonucleotide comprising at least one uridine and at least one modified uridine,

contacting said target oligonucleotide with psoralen, and

coupling said psoralen to said target oligonucleotide to form a monoadduct.

29. A method according to Claim 28, wherein said modified uridine is modified to avoid coupling with psoralen.

30. A method according to Claim 28, wherein said modified uridine is pseudouridine.

31. A method according to Claim 28, wherein said target oligonucleotide is a tRNA molecule selected from the group consisting of tRNA, modified tRNA and tRNA analogs.

32. A method according to Claim 28, wherein said target oligonucleotide is a mRNA molecule selected from the group consisting of mRNA, modified mRNA and mRNA analogs.

33. A method according to Claim 28, wherein psoralen is coupled to said target oligonucleotide by one or more cross-links.

34. A method according to Claim 28, wherein the method further comprises providing a second oligonucleotide comprising a nucleotide sequence complementary to the target oligonucleotide sequence, wherein the second oligonucleotide contains no uridine.

35. A method according to Claim 28, wherein the method further comprises providing a second oligonucleotide comprising a nucleotide sequence complementary to the target oligonucleotide sequence, wherein said second oligonucleotide comprises uridine

residues and wherein the uridine residues are modified to avoid cross-linking with the target oligonucleotide.

36. A method according to Claim 35, wherein the modified uridine is pseudouridine.

Figure 1

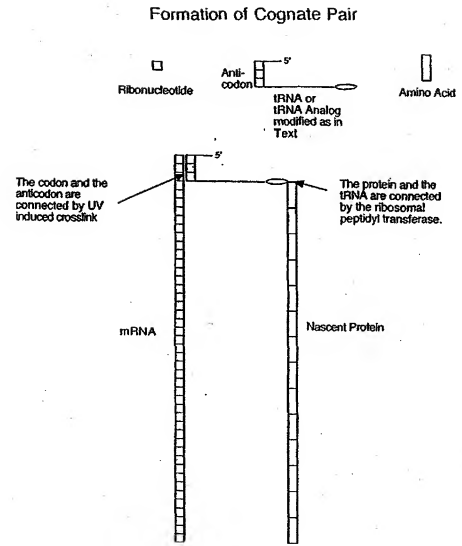

Formation of Cognate Pair
Ribonucleotide
Anti-codon
5'
Amino Acid
tRNA or tRNA Analog modified as in Text
5'
The codon and the anticodon are connected by UV induced crosslink
The protein and the tRNA are connected by the ribosomal peptidyl transferase.
mRNA
Nascent Protein

Figure 2

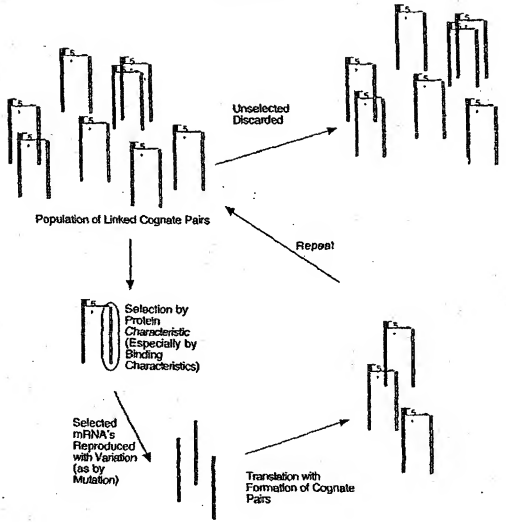

Unselected
Discarded
Population of Linked Cognate Pairs
Repeat
Selection by
Protein
Characteristic
(Especially by
Binding
Characteristics)
Selected
mRNA's
Reproduced
with Variation
(as by
Mutation)
Translation with
Formation of Cognate
Pairs

Figure 3

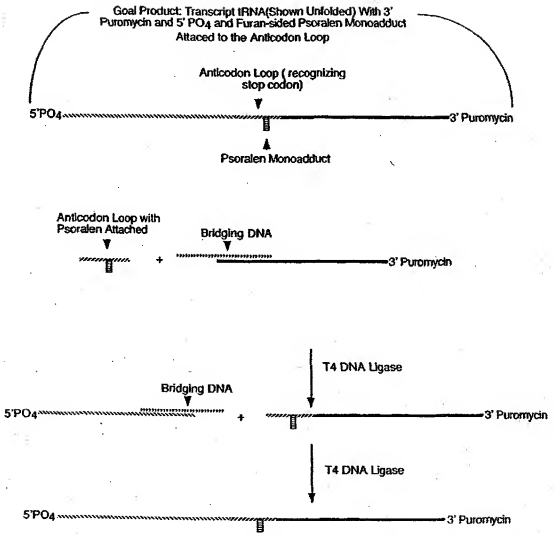

Goal Product: Transcript tRNA(Shown Unfolded) With 3' Puromycin and 5' PO4 and Furan-sided Psoralen Monoadduct Attaced to the Anticodon Loop
Anticodon Loop (recognizing stop codon)
5'PO4
3' Puromycin
Psoralen Monoadduct
Anticodon Loop with Psoralen Attached
Bridging DNA
3' Puromycin
T4 DNA Ligase
Bridging DNA
5'PO4
3' Puromycin
T4 DNA Ligase
5'PO4
3' Puromycin

Figure 4

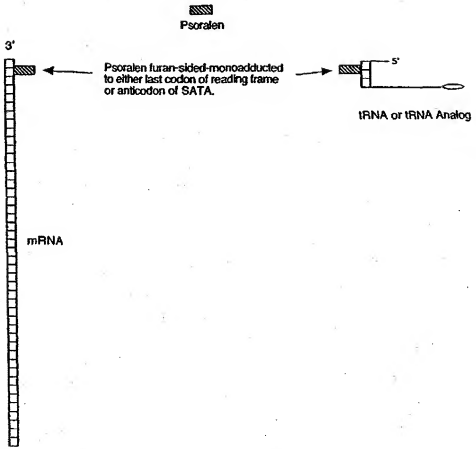
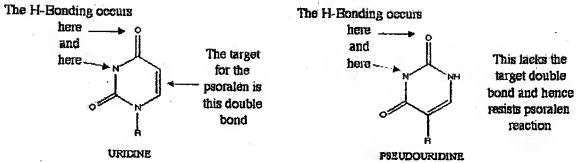

Psoralen
3'
Psoralen furan-sided-monoadducted to either last codon of reading frame or anticodon of SATA.
5'
tRNA or tRNA Analog
mRNA

Figure 5


The H-Bonding occurs
here
and
here
The target for the psoralen is this double bond
URIDINE
The H-Bonding occurs
here
and
here
This lacks the target double bond and hence resists psoralen reaction
PSEUDOURIDINE

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US02/37103

A. CLASSIFICATION OF SUBJECT MATTER
IPC(7) : C12N 15/10; C12P 21/00; C12Q 1/68; G01N 33/53,33/566; C07H 21/02; C08H 1/00
US CL : 435/6, 7.1, 69.1; 520/402; 204/157.15
According to International Patent Classification (IPC, and national classification and IPC

B. FIELDS SEARCHED
Minimum documentation searched (classification system followed by classification symbols)
U.S. : 435/6, 7.1, 69.1; 520/402; 204/157.15

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Please See Continuation Sheet

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,P	US 2002/0031762 A1 (MERRYMAN et al) 14 March 2002(14.03.2002), entire document, especially paragraphs 0010-0025.	1-4, 14-19 and 21
X,P	US 6,429,300 B1 (KURZ et al) 06 August 2002(06.08.2002), entire document, especially col. 1-col. 4.	1 and 19-21
A	US 6,261,804 B1 (SZOSTAK et al) 17 July 2001(17.07.2001), entire document..	1-22

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	
"E" earlier application or patent published on or after the international filing date	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	"&" document member of the same patent family

Date of the actual completion of the international search 10 March 2003 (10.03.2003)	Date of mailing of the international search report 10 APR 2003
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703)305-3230	Authorized officer Gromwen M. Loeb Telephone No. 703-308-0196

Form PCT/ISA/210 (second sheet) (July 1998)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US02/37103

Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claim Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☒ Claim Nos.: 23
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
The meaning of claim 23 is so illegible as to be meaningless. First, there is no antecedent basis for the phrase "said reaction to a protein identifier". Second, it is unclear what the phrase "reaction to a protein identifier" means.

3. ☐ Claim Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
Please See Continuation Sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☒ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest
☐ The additional search fees were accompanied by the applicant's protest.
☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

PCT/US02/37103

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be search, the appropriate additional search fees must be paid.

Group I, claim(s) 1-22 and 25-27, drawn to a method for selecting a desired protein or nucleic acid molecule.

Group II, claim(s) 24 and 28-36, drawn to a method of forming a monoadduct.

The inventions listed as Groups I and II do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: Group I has the special technical feature, not shared by the other group, of providing at leat two candidate mRNA molecules wherein at least one of the mRNA molecules contains at least one of a stop codon or a pseudo stop codon, translating the mRNA, linking the mRNA to its corresponding translated protein to form a cognate pair and selecting one or more cognate pair based upon properties of said translated protein or said mRNA molecule. Group II has the special technical feature, not shared by the other group, of coupling psoralen to a target oliognulcleotide comprising at least one uridine and at least one modified uridine.

Continuation of B. FIELDS SEARCHED Item 3:

MEDLINE, BIOSIS, CAPLUS, EMBASE, EAST (US Patents DATABASE)

search terms: tRNA, mRNA, protein or polypeptide, crosslink, psoralen, modified uridine or pseudouridine, monoadduct, puromycin, RNA ligat? protein or peptide or polypeptide